Artificial Intelligence and Human-Computer Interaction
Y. Ye and P. Siarry (Eds.)

doi:10.3233/FAIA240150

The Composite Multi-Level Prediction Model for Stock Market Time Series

Zhichao CHANG[1] and Zuping ZHANG
School of Computer Science and engineer, Central South University, China
ORCiD ID: Zhichao Chang https://orcid.org/0000-0002-6253-242X

Abstract. For a long time, investors have always been influenced by their own experience and investment expert advice. Machine Learning method in Quantitative Investment is an advanced method that replaces subjective judgments with artificial intelligence models to improve transaction accuracy. The authors have implemented a composite stock price prediction model based on multi-layer training networks, which is more suitable for predicting future stock prices compared to traditional methods. This model starts from the time series of stock data and deeply integrates the characteristics of stocks with artificial intelligence. The deep training sub-model is a combination of machine learning models and traditional statistical methods. This unique design cleverly solves the one-sidedness of machine learning methods and the petrification of classical financial methods. Through comparative experiments on multiple prediction models within the same time period, the model proposed in this paper has been proven to be the most effective model.

Keywords. Stock market, machine learning, pre-training model, deep training model, historical data set

1. Introduction

A stock is a voucher provided by a stock company to the purchaser for sharing benefits and assuming obligations. It provides assistance for the company to raise social funds. Stock trading can be seen by investors as a means of profitability, as well as by financial researchers as a display window of the country's economic development status. In order to grasp the trend of stock trading, participants need to develop effective investment strategies. Therefore, people hope to start with historical data to obtain effective information and develop accurate investment strategies. Historical stock data contains information about the past trends of stocks and may reveal the possible upward and downward directions of stocks in the future. By analyzing the historical data, researchers can use a series of calculations to describe the possible future trends of the stock market. Through these means, researchers can provide investors with key information on when to purchase which stocks.

In recent years, the use of machine learning in the financial services industry to replace manual decision-making has always been a focus of media. Machine learning methods are frequently used in the financial field, such as loan decision-making, investment strategies, and company qualification judgment. (See Wall 2018)

[1] Corresponding Author: Zhichao Chang, Email: changzhichao1@mail.csu.edu.cn

With the widespread use of computer devices in global stock markets, massive amounts of stock market data can be continuously obtained from company reports, forums, and transaction record terminals. As valuable assets for various enterprises, these massive amounts of data can bring a lot of additional benefits to the enterprise. However, historical data of the stock market is a type of time series data, which has characteristics of nonlinearity and high latitude [1]. Moreover, the results of a single transaction are influenced by both the previous transaction and the subsequent transaction results. For time series $t_1<t_2...<t_n$, if the data value t_2 at a certain point is changed by the outside world, it will also affect the data values of adjacent t_1 and t_3 time points at that point. The complexity of stock market historical data makes the efficiency of ordinary financial statistical methods invalid. The authors attempt to use nonlinear deep machine learning models to extract and analyze key information. Although the efficient market theory suggests that no method can obtain additional returns from the stock market, many researchers believe that this can be achieved through reasonable methods [2-4]. Therefore, the author of this article made a reasonable prediction of the stock price trend based on this premise.

Due to the increase in stock market trading volume, the data processing problem has become increasingly inconvenient to handle. Therefore, the industry has widely adopted computer processing, also known as deep learning models, to process massive datasets [5-7]. According to the Massachusetts Institute of Technology Review, Goldman Sachs Group replaced human traders with computers, reducing their number from 600 to 2. The reason for this phenomenon is that deep learning methods can mine nonlinear and high-dimensional information from historical data, which cannot be achieved by econometric models and expert analysis. Therefore, in the financial field, more and more participants have joined the research on deep learning models and more efficient artificial intelligence tools have replaced the work of ordinary data analysts.

The contributions of this paper are:

- This article proposes a multi-level composite prediction model, with each level completing different functions. The pre-training layer of this model can identify different price categories in the total sample through attention mechanism. The deep training layer can deeply mine information from various categories and predict the sample distribution that best matches the true stock value.
- The model proposed in this article uses both machine learning and classical financial methods during the training phase, making the prediction results more flexible and regular. The method introduced in this paper achieves an average accuracy of over 90% in predicting stock prices of companies.
- The model proposed in this article uses both machine learning and classical financial methods during the training phase, making the prediction results more flexible and regular. The method introduced in this paper achieves an average accuracy of over 90% in predicting stock prices of companies.

This article consists of six parts. Chapter 2 introduces the existing models used for stock price prediction and their advantages and disadvantages compared to the methods introduced in this paper. Chapter 3 introduces the basic theories and related methods used in the model mentioned in this article. Chapter 4 is an introduction to the details of the experiment. This section includes an introduction to the dataset, the methods proposed in this article, and the initial work used in other methods. Chapter 5 describes the comparison and presentation of prediction results generated using various methods. Chapter 6 is the summary section, which summarizes the ideas of the paper.

2. Related literature

Except in the stock market, time series prediction plays an important role in many research fields. However, the research object has the characteristics of strong interference and high chaos. This makes the task of making predictions for complex and challenging time series a daunting challenge.

For a long time, investors generally used classic financial methods to manage their funds [8-10]. The buy and hold method is an ancient method that assumes that each stock fluctuates around its true value. Therefore, investors need to hold the shares for a long time after buying them until the end of the transaction. The mean reversion method believes that no matter how the market price fluctuates, the stock value will return to the average value of the market. The time series Momentum method uses the idea of objects maintaining inertial motion in physics, believing that the fluctuation of stock prices also conforms to the principle of the strong maintaining strength. The conclusion drawn from this method is that the price trend of strong stocks is also strong.

In recent years, due to the surge in financial data volume and the need for convenient processing, quantitative trading has been fully utilized. Thanks to cheaper computing resources and easier access to data, researchers are gradually using machine learning methods - typically neural networks to replace traditional techniques - typically logistic regression for predicting defaults. As the former reduces work costs while ensuring efficiency does not decrease. Currently, many mature machine learning methods such as Transformer [11], Reinforcement Learning [12-13] and DeepScalper [14] are used for stock predictions. Chen [15] believe that the trend of each stock is closely related to other stocks in the same market. Therefore, he uses machine learning methods to convert multiple stocks into vectors and quantifies the relationships between each stock through formulas, which serve as the basis for determining the trend of a single stock. In [16], The authors use natural language processing technology to process short and long comment articles and accurately predict the one-day trend of stock prices.

Whether using machine learning methods or classical financial methods, their results are only effective for some time periods. Due to the complexity of stock data, both methods have some drawbacks in producing results. Because the former focuses on the overall distribution of data and ignores the regularity of true distribution. The latter focuses on the simple regularity of distribution, neglecting the integrity of the data. The model proposed by authors of this paper combine these two methods to overcome their one-sidedness and rigidity. In order to comprehensively analyze the high/low latitude information of stock data and fully improve prediction efficiency, we describes a hybrid prediction system. The system includes a pre-trained model for rough classification and a deep training model for deep operations that includes multiple responsibilities.

3. Theoretical preparation

3.1. Normalization

Data normalization processing is a fundamental task in financial data processing. Different attribute variables often have different value ranges, and the differences between different features may be significant. If we do not process it, it may affect the results of data analysis. To eliminate order of magnitude differences between indicators, we need to call the maximum-minimum standardization formula for standardization processing. This formula used in this paper is as follows:

$$X_{nom} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

$$X = X_{nom}(X_{max} - X_{min}) + X_{min} \tag{2}$$

where X_{min} / X_{max} is the minimum / maximum of each attribute variable dataset. X is a initial value. Eq(1) is used to convert from an initial value to a standard value X_{nom} which is between 0 and 1, and the result won't have any bias when data is processed by neural network. Eq(2) is used to restore Xnom to its initial value X.

3.2. Introduction for GAN and Wasserstein GAN

The training process of Generator adversarial networks(GAN) usually includes the following steps:

- Generator network generates noise data and inputs it into the discriminator network for judgment.
- Discriminator network classifies input data and calculates classification errors.
- Update the parameters of the discriminator network based on classification errors to improve its classification accuracy.
- Generate a network that updates its own parameters based on the feedback information of the discriminative network to generate more realistic simulation data.
- Repeat the above steps until the preset convergence conditions are reached.

Wasserstein GAN(WGAN) [17] is an alternative to traditional GAN training. WGAN can improve the stability of learning, get rid of problems such as pattern collapse, and provide a meaningful learning curve useful for debugging and hyperparameter search. The generator of WGAN cell used in this paper consists of three layers: convolution and pooling layer, full connection layer and activation function layer. Convolution can reduce data dimensions and remove redundant information. Pooling is used to prevent overfitting. The fully connected layer can map the original data from the current working layer and map the learned "distributed feature representation" to the sample label space. The activation function layer includes nonlinear mapping, which can approximate any nonlinear function to improve the expression ability of the entire neural network. The loss function of WGAN is:

$$W(G) = \max_{w \in W} E_{x \sim p_r} | f_w(x) | - E_{z \sim p_z} | f_w(g_\theta(Z)) | \tag{3}$$

where x is a subset of real data, and Z is a subset of the true subset. By training the weight w, the model updates the backpropagation algorithm, and then makes the results generated by the generator closer to the values generated by the discriminator. To achieve the best performance of generator G, the effect of reducing the gradient must be just right.

3.3. ARIMA-LASSO unit

Although some researchers have achieved certain results in using the ARIMA model to predict and process temporal data, the ARIMA model only considers the impact of time factors on the results and cannot reflect the internal relations between many influencing factors that cause the results. For example, experts in the financial field [18] conducted long-term empirical analysis of stock market data based on various theoretical methods such as econometrics and statistics, and used ARIMA to improve the accuracy of stock price prediction. If we combined ARIMA with another algorithm, LASSO regression, for temporal data prediction, this problem can be effectively solved. By adding a penalty term to the regression model, LASSO regression can compress too small regression coefficients to 0 and eliminate them. Then, the ARIMA-LASSO unit implemented automatic screening of independent variables, thereby achieving model simplification while ensuring model stability.

3.4. BIGRU and attention lay

Many prediction models have the problem of long training time and insufficient feature extraction. To solve this problem, the authors [19] adopt a hybrid model of Bidirectional Recurrent Neural Network (BIGRU) and attention mechanism. BIGRU has the ability to prestore much data and process it in both directions, allowing for faster data processing. The attention mechanism can capture multiple features over long distances, improving the robustness of processing irrelevant information.

4. Theoretical preparation

Although models such as GRU can be used to analyze stock data, ordinary AI models cannot overcome the impact of low latitude information. So the authors consider the particularity of stock data and collaborate with the dataset for two training sessions. In the pre-trained model, the system can obtain basic data classification information with only a small amount of training. Deep training models can obtain more accurate results by training datasets based on classification. This process separates common information and special information, greatly improving the training efficiency of the system. The system proposed by the author called Multivariate Learning Stock price prediction system(LM-SPP). The overview of the proposed LM-SPP framework is shown in Figure 1.

4.1. BIGRU and attention lay

The pre-trained model includes a bidirectional GRU layer, a self attention layer, an MLP layer, and a comparison layer. In the pre-training layer, the input data is connected to the previous/next processing results through the caching mechanism provided by BIGRU to generate new results. Next, the attention mechanism layer extracts high latitude information from low dimensional data. After that, the samples are arranged in descending order according to the generated predicted values. Finally, the comparison layer generated three classification sample sets by comparing the difference between the upper /lower threshold values and the resulting predicted stock values. The initial samples contained in each section were not robust. We use the Z-test method to test the stability of samples in each class and dynamically move samples from unsuitable categories to suitable ones. When the Z-test results in all classes are correct, we can obtain three completely robust stock price classifications. After classification, the variance in each class is the smallest. As shown in Figure 1(A), these three types of sample sets represent different investment emotions: The buying set represents the system's optimism towards the stocks of the day and can be accessed; The selling set symbolizes that the system is not optimistic about the stocks of the day and needs to sell; The holding set depicts the system's belief that it is currently in the closing stage and needs to be held.

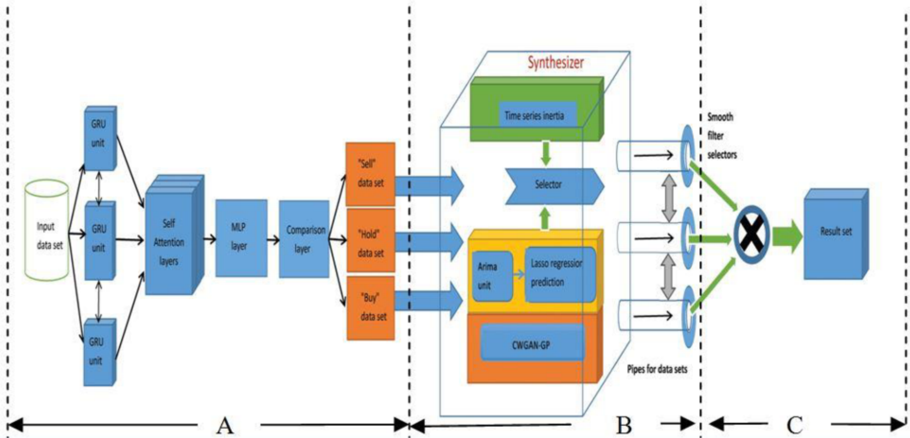

Figure 1. An overview of the proposed LM-SPP framework. We show three individual building blocks: (a) BIGRU and ARIMA-LASSO lay, (b) Synthesizer, (c) Smooth filter selectors.

As shown in Figure 1(B), the deep training network is a composite stock price generation module that includes a synthesizer and a smoothing filter. The synthesizer consists of a generative adversarial network, ARIMA-LASSO unit, Time series inertia and a selector. In the core module of prediction value generation, ARIMA-LASSO units will reconcile data with significant prediction errors generated by generative adversarial networks. The purpose of designing this module is to overcome the problem of interference from different types of data, as well as the problem of result distortion caused by complete reliance on the machine judgment.

When the predicted results fluctuate significantly, the error in the results will be significant. After a period of rapid increase or decrease in stock prices, there is a possibility of significant directional trend changes in the following days. We call this phenomenon as the consolidation stage. This stage is generally longer than 5 days, mainly due to: 1). The vast majority of retail investors exhibit a "herd" mentality, lacking resilience and courage to invest. 2). The main force of shareholding will adjust the highly volatile stock price, allowing investors who are chasing gains and profits to voluntarily surrender their chips. Therefore, this article uses the formula for the angle between two vectors to determine whether the prediction results for a certain period have deviated too much. The formula for calculating the angle β is as follows:

$$\beta = \arccos\frac{AB}{|A||B|} = \arccos\left(\frac{\partial(PS_{N+\delta} - PS_N).\partial(PS_{N+2\delta} - PS_{N+\delta})}{\sqrt{(PS_{N+\delta} - PS_N)^2 + \delta^2} * \sqrt{(PS_{N+2\delta} - PS_{N+\delta})^2 + \delta^2}}\right) \tag{4}$$

where $PS_N, PS_{N+\delta}, PS_{N+2\delta}$ is the predicted stock value on day N,N+δ,N+2δ. A is a vector:PS_N->$PS_{N+\delta}$, B is a vector:$PS_{N+\delta}$->$PS_{N+2\delta}$. When the angle between A and B is large, it indicates that the stock market fluctuates significantly after the nth day. System assumption: When the inner angle β>75° and the consolidation time is 6 days, the fluctuation of the predicted value is too large. The selector will call the time series inertia to generate corrected predicted values when the stock price fluctuates significantly.

In order to address the issue of mutual interference caused by the overlapping dates of different data sources, we need to correct the result PS_N generated by the generated adversarial network on day N using the following formula:

$$PS_N = TS_N(\beta > 75^\circ, T_{N-1} - PS_N > T_{N-1} - TS_N) \tag{5}$$

where TS_N is the value generated by the time series inertia, and T_{N-1} is the true stock value on day N-1.

As shown in Figure 1(C), different categories will be output through different pipelines on the synthesizer, and the corresponding smoothing selector for each pipeline will correct result that deviate significantly from the true trend curve. The formula is as follows:

$$PS_j = \frac{\sum_{j=1}^{\frac{R}{2}} PS_{\frac{R}{2}-j} + \sum_{j=1}^{\frac{R}{2}} PS_{\frac{R}{2}+j}}{R} \tag{6}$$

where PS_j is the data that needs to be modified within the abnormal range S. The expectation of this method is to update the abnormal data based on the left and right values of the abnormal data. R is the number of elements within the abnormal range S.

5. Experiment

By collecting 3-year historical data from 3 companies in the Shanghai stock market of China, we divided the historical data of these 3 companies into training and testing sets. The training set consists of data from one and a half year, and the testing set consists of historical data in the closest six months to now. These are special stock market data sets, which contain 11 cause characteristics (independent variables) and a result label (dependent variable). The reason why the closing price is used as a label is that the closing price can reflect the degree of attention of market funds to a certain stock and indicate the direction of the trend of the next trading day. The closing price has great guiding significance for short-term investors, especially in the tail market. If the stock price closes at a higher price on the first day, it will often have a more prominent performance on the next day.

The prediction process is summarized as follows:

- In the pre-training layer, according to eq(1) and eq(2), the stock price dataset is divided into three categories (go up, go down, hold) based on trends. 90% of the total dataset is divided into training sets, and 10% of the data is divided into testing sets.
- In the deep level model, the system does not particularly distinguish between past information and future information. The data of each category is deeply trained in the WGAN model, and the resulting results are rearranged in chronological order. This generates predicted values that simulate the distribution of real data. The distribution of this predicted data shows a relatively gentle upward and downward trend. Therefore, except for some singular points, the predicted data will be very similar to real data.

In order to test the predicted effect of using different models to predict the stock price of each company, Table 1 is shown below. We analyzed the different performances of 3 stocks under multiple models. Table 1 shows the MSE, MAPE, and MAE of test data under different method models. We can get the following points from the Table 1. The prediction results of the last three methods mentioned in this manuscript are significantly superior to other commonly used methods. 2. Except for the obvious inapplicability of AT-GRU, other methods can be used as alternative methods for stock price prediction. 3. LM-SPP improves the predicted effect of WGAN and ARIMA-Lasso, so it is the most suitable method.

Table 1. Comparison of prediction effects of various models.

Code of different companies	Models	MSE	MAPE	MAE
Three's Company (SH605168)	AT-GRU	0.0591	0.2623	0.1911
	DeepScalper	0.0955	0.3064	0.1030
	ARIMA-Lasso	0.0636	0.3953	0.2118
	WGAN	0.0648	0.2244	0.1324
	LM-SPP	0.0245	0.1535	0.0843
SUPCON (SH300097)	AT-GRU	0.1521	0.3952	0.2535
	DeepScalper	0.0475	0.2037	0.3532
	ARIMA-Lasso	0.0297	0.2877	0.0773
	WGAN	0.0286	0.1856	0.0489
	LM-SPP	0.0284	0.1722	0.0418
Jinhe Business Management (SH603682)	AT-GRU	0.5100	1.2150	0.6748
	DeepScalper	0.1650	0.3941	0.2184
	ARIMA-Lasso	0.1785	0.2568	0.1176
	WGAN	0.1236	0.2544	0.1121
	LM-SPP	0.0887	0.0495	0.0217

Through comparative experiments, it can be seen that LM-SPP is similar to WGAN when measuring the difference between the predicted distribution and the real distribution. They all hope to improve the accuracy by controlling the linear slope of the elements in the prediction distribution and the elements in the target distribution. The method we recommend will fine-tune the predicted values generated by multiple generators at the same time to make the results more reasonable and accurate.

Moreover, the author considers using a line chart to demonstrate the predicted effect of different prediction models. When dealing with the same stock, there are always some models that perform well, and others that perform poorly. In Figure 2, the effects of multiple prediction models in the same time period are demonstrated respectively. Figure 2 shows the closing price trend of Three's Company (code: SH605168) from January, 2023 to June, 2023. We can find that all the distributions have many deviations from the original distribution except LM-SPP. In addition to using the LM-SPP method, the fluctuations of other distributions are large. All experimental results show that WGAN is superior to other methods except LM-SPP on the stock data set of Three's Company. Therefore, when forecasting the real stock price, we should give priority to LM-SPP and WGAN's improved methods.

The method introduced in this article achieves average accuracy of over 90% in predicting stock prices of companies. However, the LM-SPP model has lower prediction accuracy compared to some classic prediction models such as SVM, LSTM, RF, etc. Because these models typically use the previous day's closing price to predict today's stock value. Although it may seem that the accuracy of predictions are high, they are generally an imitation procession of the past stock prices, not some methods that conform to financial logic. The LM-SPP model is formed by absorbing the advantages of other time series prediction methods that have been proven by other paper. It is an evolutionary result of stock price prediction. It has been proven that the LM-SPP model can predict the vast majority of stock prices on the market. Therefore, we can conclude that the LM-SPP model is an excellent model that inherits the advantages of past reasonable models.

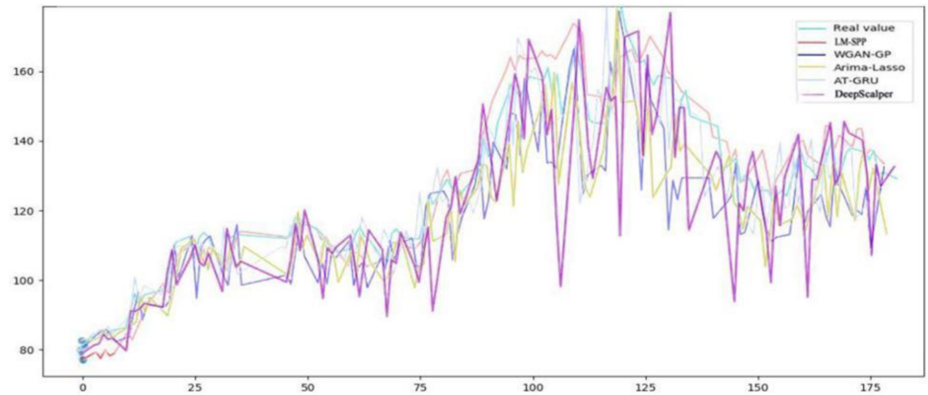

Figure 2. The effects of multiple prediction models in the same time period for Three's Company.

6. Conclusion

In this paper, the author considers processing the classification generated by the pre-training network separately in each generator and synthesizes the processing results. Finally, generators imitate new distribution that approximates the real distribution. Unlike traditional WGAN structures, the generators of LM-SPP cannot completely determine the prediction results. Through the cooperation of the scorer, regulator, and generator, the most realistic imitation data is generated. The final result is to obtain the value of the predicted stock efficiently.

References

[1] Brock WA. Nonlinearity and complex dynamics in economics and finance[J]. Economic Development & Cultural Change. 1988; 53(3):760-762, doi: 10.1086/ 426417.

[2] Li L, Matt P A, Heumann C. Forecasting foreign exchange rates with regression networks tuned by Bayesian optimization[J]. Papers. 2022. doi:10.48550/arXiv.22 04.12914.

[3] Luo Z, Guo W, Liu Q, et al. A hybrid model for financial time-series forecasting based on mixed methodologies[J]. Expert Systems. 2020;38(1), doi:10.1111/exsy.12633.

[4] Caldera H, Lavanya W. Combinatorial Impact of Technical Indicators on Price Prediction in Colombo Stock Market. 20th International Conference on Advances in ICT for Emerging Regions. 2020:256-261, doi: 10.1109/ICTer51097.2020.932 5500.

[5] Wang F, Li M, Mei Y, et al. Time Series Data Mining: A Case Study with Big Data Analytics Approach[J]. IEEE Access. 2020; PP (99):1-1, doi:10.1109/ACCESS.2020.2966553.

[6] He H, Gao S, Jin T, et al. A hybrid model for financial time-series forecasting based on mixed methodologies[J]. Applied Soft Computing. 2021;108(5952):107488. doi: 10.1016/j.asoc.2021.107488.

[7] Wiese M, Knobloch R, Korn R, et al. Quant GANs: deep generation of financial time series[J]. Quantitative Finance. 2020, 20. doi:10.1080/14697688.2020.1730 426.

[8] Drenovak M, Rankovi V, Uroevi B, et al. Mean-Maximum Drawdown Optimization of Buy-and-Hold Portfolios Using a Multi-objective Evolutionary Algorithm[J]. Finance Research Letters. 2021(1): 102328. doi: 10.1016/j.frl.2021.102328.

[9] Nartea G V, Valera H G A, Valera M L G. Mean reversion in Asia-Pacific stock prices: New evidence from quantile unit root tests[J]. International Review of Economics & Finance. 2021. doi: 10.1016/j.iref.2020.12.038.

[10] Boubaker S, Liu Z, Lu S, et al. Trading Signal, Functional Data Analysis and Time Series Momentum[J]. Finance Research Letters. 2021, 42(3): 101933. doi: 10.1016/j.frl.2021.101933.

[11] Ding Q, Wu S, Sun H, et al. Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction[C]. International Joint Conference on Artificial Intelligence. 2020. doi: 10.24963/IJCAI.2020/640.

[12] Fang Y, Ren K, Liu W, et al. Universal Trading for Order Execution with Oracle Policy Distillation[J]. In Proceedings of the 35th AAAI Conference on Artiffcial Intelligence. 2021:107–115. doi:10.48550/arXiv.2103.10860.

[13] Liu Y, Liu Q, Zhao H, et al. Adaptive Quantitative Trading: An Imitative Deep Reinforcement Learning Approach[J]. Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(2): 2128-2135.doi:10.1609/aaai.v34i02.5587.

[14] Sun S, Wang R, He X, et al. DeepScalper: A Risk-Aware Deep Reinforcement Learning Framework for Intraday Trading with Micro-level Market Embedding[J]. Papers. 2021.doi:10.48550/arXiv.2201.09058.

[15] Chen W, Jiang M, Zhang W G, et al. A novel graph convolutional feature based convolutional neural network for stock trend prediction[J]. Information Sciences, 2020, 556(3). doi:10.1016/j.ins.2020.12.068.

[16] Chang Z, Zhang Z. Judging Stock Trends According to the Sentiments of Stock Comments in Expert Forums. Electronics. 2023, 3:722. doi: 10.3390/electronics 12030722.

[17] Cintas C, Speakman S, Tadesse G A, et al. Pattern detection in the activation space for identifying synthesized content[J]. Pattern Recognition Letters, 2022, 153. doi: 10.1016/j.patrec.2021.12.007.

[18] Ding Y, Sun N, Xu J, et al. Research on Shanghai Stock Exchange 50 Index Forecast Based on Deep Learning[J]. Mathematical Problems in Engineering. 2022.doi: 10.1155/2022/1367920.

[19] Niu D, Yu M, Sun L, et al. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism[J]. Applied Energy. 2022, 313:118801-. doi: 10.1016/j.apenergy.2022.118801.